

World Bank Lending and the Quality of Economic Policy

Lodewijk Smets^{a,b,*}, Stephen Knack^c

^a*Institute of Development Policy and Management, University of Antwerp, Belgium*

^b*LICOS Centre for Institutions and Economic Performance, University of Leuven, Belgium*

^c*World Bank, Washington DC*

Abstract

This study investigates the impact of World Bank development policy lending on the quality of economic policy. It finds that the quality of policy increases, but at a diminishing rate, with the cumulative number of policy loans. Similar results hold for the cumulative number of conditions attached to policy loans, although quadratic specifications indicate that additional conditions may even reduce the quality of policy beyond some point. The paper measures the quality of economic policy using the World Bank's Country Policy and Institutional Assessments of macro, debt, fiscal and structural policies, and considers only policy loans targeted at improvements in those areas. Previous studies finding weaker effects of policy lending on macro stability have failed to distinguish loans primarily intended to improve economic policy from other loans targeted at improvements in sector policies or in public management. The paper also shows that investing in economic policy does not “crowd out” policy improvements in other areas such as public sector governance or human development. The results are robust to using alternative indicators of policy quality and correcting for endogeneity with system generalized methods of moments and cross-sectional two-stage least squares. The more positive results in the study relative to some previous studies are consistent with claims by the World Bank that it has learned from its mistakes with traditional adjustment lending.

Keywords: development policy lending, Economic policy, Aid effectiveness, World Bank

1. Introduction

Since 1980 the World Bank has been providing conditional financing to recipient governments to support specific policy and institutional reforms. These development policy loans (DPLs) – formerly known as structural adjustment lending (SAL) – have become an important component in the financing of development operations. For instance, in fiscal year 2008 they accounted for 6.6 billion USD or 27 percent of total World Bank commitments.

Not surprisingly, there exists a vast literature evaluating the effects of adjustment lending. However, no clear consensus view emerges from this research as some studies find a positive effect of adjustment lending on growth and macroeconomic policies, while others indicate that policy lending failed to induce change with no significant impact on growth.

The lack of consensus is in part due to methodological challenges encountered in examining the effectiveness of policy lending. This study investigates the impact of World Bank lending on the quality of policy, addressing three particular methodological concerns.

First, there is a potential selection bias problem. Countries often receive policy loans because of policy deficiencies, so the coefficient on policy lending may be biased downward when examining its impact on policy outcomes (Easterly, 2005). On the other hand, the coefficient may be biased upward, if loans tend to go to motivated governments that would have reformed even in the absence of support. Hence, estimating the impact of development policy lending calls for a robust identification strategy, which we implement with instrumental variable estimation and system GMM.

Second, it is important to select appropriate dependent variables. World Bank loans seek to improve policy in many different sectors or sub-sectors (see the corresponding table in the online appendix of this journal), and the estimated impacts of lending may be biased

*Corresponding author

Email addresses: `lode.smets@kuleuven.be` (Lodewijk Smets), `sknack@worldbank.org` (Stephen Knack)

downward if the outcome variable is not matched with the relevant subset of policy loans. In contrast with much of the existing literature on DPL effectiveness, we adjust for the policy target of World Bank lending. For example, Easterly (2005) acknowledges that his study is limited to ‘easily quantifiable [objective] macroeconomic indicators’ and that DPLs also target other policy improvements, such as reform of inefficient financial sectors.

Third, as theory provides little insight on how development policy lending affects policy quality, we also examine potential scale effects. Specifically, we test different functional forms that allow for increasing or decreasing returns to additional loans (or conditions).

Another possible explanation for the divergent findings in the literature is the time period under investigation. Most studies evaluate the first two decades of adjustment lending. At that time, the contracts offered implied a policy of ex-ante, donor-driven lending.¹ Given the shortcomings of this approach, the World Bank modified its policy towards adjustment lending around the turn of the millennium. The more positive results of the few (internal) reviews evaluating recent episodes of adjustment lending could indicate an improved effectiveness of policy support. However, as a robust econometric study is still lacking, this paper aims to fill this gap by investigating the period 1995-2008.

Results from panel estimations show that the number of DPLs has a positive but diminishing effect on the quality of economic policy. This finding is robust to sample restrictions, additional controls, the use of alternative indicators of policy quality, and correction for endogeneity with system GMM. Further evidence is provided by instrumenting our variable of interest – the number of cumulative economic policy loans – in a cross-sectional setting. Similar results are obtained when we substitute the number of cumulative conditions for the number of DPLs as the key regressor, although here it is less clear which functional form best fits the data.

We further test whether implementation of economic policy loans “crowds out” policy improvements in other, non-targeted policy areas. Conceivably, improving policy in one

sector or sub-sector might divert rent-seeking efforts to other sectors. However, we find no evidence in our tests that investing in economic policy significantly affects policy quality in other areas such as public sector governance or social sector and environmental policies.

The remainder of the paper is structured as follows. In the next section we present a brief history of World Bank policy lending and review the related literature. Section 3 describes the data and methodological issues. Section 4 presents the empirical results. In that section, we first discuss findings from the panel estimations using the number of cumulative loans and the number of cumulative conditions as key variables of interest. For both variables, we test linear, quadratic and logarithmic model specifications. Next, we show that our main results are robust to sample restrictions, additional controls and the use of alternative indicators of policy quality. In subsection 4.3, we address endogeneity concerns and discuss the results from system GMM and cross-sectional 2SLS. Finally, section 5 concludes.

2. Background

In 1980 the World Bank launched its first non-project lending instrument to support policy change in recipient countries. At that time, top management was dissatisfied with the limited influence of the Bank's normal project lending on policies of borrowing governments. Therefore structural adjustment lending was conceived, as a new lending program with which the Bank would try to help countries to tackle important policy deficiencies. The programs provided conditional finance in support of specific policy reforms. In its early years adjustment lending mainly emphasized economic stabilization and correction of balance of payments distortions. At the beginning of the 1990s more emphasis was put on protecting the poor from the adverse effects of the adjustment programs. The contracts that were offered implied a policy of ex-ante, donor-driven lending (Kapur et al., 1997).

However, as the introduction of structural adjustment lending (SAL) generated concerns

from within the Bank and from borrowing countries (World Bank, 1989),² several studies investigated its effectiveness. Internal World Bank reviews indicated that early adjustment lending produced mixed results. For instance, comparing program with non-program countries in a before-after analysis, World Bank (1989) found that policy lending stimulated growth and balance of payments performance. Interestingly, results of this exercise were more favorable when intensive program countries – that is, countries that received three or more adjustment loans – were compared with non-program countries. However, the study also noted that target countries had not been able to grow out of debt (as envisioned) and questioned the sustainability of reforms. Taking a sectoral approach, Jayarajah and Branson (1995) analyzed the effectiveness of SAL using evaluation audits and project completion reports for 99 adjustment operations, covering the period 1980-1992. Again, mixed results were found; for example, only 24 of the 40 countries that received macroeconomic adjustment loans were able to reduce fiscal deficits and bring down inflation.

In addition to those internal evaluations, external research also examined the performance of adjustment lending. Two early studies include Mosley et al. (1991) and Killick et al. (1998). Using various methodologies – comparing program and non-program countries, regression analysis and model simulations – Mosley et al. (1991) found that development policy operations were instrumental in strengthening export and balance of payments performance, but had little impact on economic growth. The authors also found that adjustment programs were associated with reduced investment. Based on a review of the literature, Killick et al. (1998) provide further evidence that early adjustment lending produced mixed outcomes. More recent studies corroborating this conclusion include Bird and Rowlands (2001), Butkiewicz and Yanikkaya (2005), Easterly (2005) and Agostino (2008). Bird and Rowlands (2001) investigate whether World Bank policy lending serves as a (positive) signal to lenders and investors. The authors attempt to correct for endogeneity by employing lagged values of their main independent variables. Using a panel of

93 developing countries that runs from 1984 to 1995, they fail to find any consistent positive effect of adjustment lending on other financial flows such as FDI, portfolio or private debt. Butkiewicz and Yanikkaya (2005) use several regression techniques to estimate the effect of World Bank adjustment lending on long-run GDP per capita growth for the period 1970-1999, correcting for endogeneity using lagged values and employing 3SLS. They conclude that World Bank lending stimulates growth in some instances, particularly in low income countries and poor democracies. In an influential paper, Easterly (2005) considers the repetition of adjustment lending to the same country as a means of reducing the selection bias problem. The author estimates a pooled probit regression over the period 1980-1999 with an extreme macroeconomic imbalance indicator as his dependent variable. Results fail to show any consistent positive effect of adjustment lending on macroeconomic stability. Additionally, Easterly (2005) examines the effect of repeated lending on growth in a cross-sectional 2SLS regression, but, again, without any significant results. Finally, based on the Heckman (1979) selection model, Agostino (2008) investigates whether signing a loan agreement has an impact on private investment. Covering the period 1982-1999, the author finds that entering into SAL has a negative effect on investment.

The mixed track record of early adjustment lending can be attributed at least in part to the limited enforceability of reform conditions (see, for example, Svensson, 2000, 2003). That is, when contracting for policy reform an independent arbitrator – an international court of law – is lacking to punish any player who breaks contract stipulations. If a recipient government cannot commit to contract conditions, the incentives provided in the (ex-ante) contract will no longer guarantee effective policy reform. A second reason for the mixed performance of SAL is poor program design and ill-chosen policies (Killick et al., 1998; Rodrik, 1990, 2008).³ For instance, Rodrik (1990) argues that a focus on liberalization is misguided if macroeconomic stability would thereby be endangered. A third reason mentioned in the literature is limited sustainability and backsliding of reforms

after implementation (World Bank, 1989; Rodrik, 1992; Collier et al., 1997). For example, World Bank (1989) indicates that many highly indebted African countries failed to maintain fiscal discipline after initial reductions in budget deficits.

Recognizing the limitations of traditional policy-based support, the World Bank modified its approach towards adjustment lending (and development assistance) around the turn of the millennium.⁴ Among other changes, it reduced the average number of conditions in its loans, strengthened country “ownership” of lending programs by using countries’ own development strategies to identify loan conditions, and moved from ex-ante towards ex-post disbursement of loan tranches (Koeberle, 2003; World Bank, 2004, 2006).⁵

Surprisingly, and in contrast to the extensive research evaluating the first two decades of adjustment lending, there is not much systematic research investigating more recent episodes of policy based lending. We found only a few internal reviews.⁶ The World Bank’s 2003 Annual Review of Development Effectiveness was dedicated to analyzing the effectiveness of Bank support for policy reform. Focusing on the period 1999-2003, the study concluded that ‘Bank lending was concentrated in countries that were improving their policies’ and that ‘in many cases’ DPLs and other Bank support ‘contributed to policy improvements’ (World Bank, 2004). Also, beginning in 2006 the World Bank provides a three-yearly retrospective of its experience with the implementation of DPLs. Overall, DPLs are evaluated favorably. For instance, comparing results to objectives, the 2009 DPL retrospective argues that DPLs have consistently achieved development outcomes during the period 2006-2009 (World Bank, 2009). Finally, a review of Bank support in fragile and conflict-affected states reports a positive and statistically significant correlation between policy improvements and the number of years under DPL support (IEG, 2013).

However, a quantitative study with a more robust identification strategy is still lacking. We aim to fill this gap by investigating the association of repeated policy lending with the quality of policy, covering the period 1995-2008. Following Easterly (2005), we focus on

repeated lending since we believe supporting policy reform is a multistage and long term process (see, for example, Pritchett and de Weijer, 2010). Our dependent variable is not a final outcome measure such as economic growth or FDI, but rather policy quality. In this choice, we are guided by Roodman (2007), who argues that development aid is probably only a weak signal in the noisy and limited data available on economic growth in developing countries. Rather than testing directly for effects on growth, we test for whether World Bank country teams achieve their objective in designing DPLs of improving the quality of development policies. In this respect our study is related to Boockmann and Dreher (2003) and Kilby (2005), who both investigate the impact of World Bank lending on the policies developing countries select, as measured by the Fraser Institute’s “Economic Freedom” index or components thereof. However, our study differs in a number of important ways from Boockmann and Dreher (2003) and Kilby (2005). For instance, Boockmann and Dreher (2003) investigate the impact of World Bank aid and IMF credits over the period 1970-1997, while our study investigates a more recent time frame. In their research, Boockmann and Dreher (2003) use a broader set of loans, including projects as well as adjustment loans that might be targeted at other policies. Furthermore, they do not explicitly address endogeneity concerns, although they do report some results using GMM. Similarly, Kilby (2005) looks at an early period, 1970-2000, and estimates the (potentially) endogenous effect of World Bank lending and aggregated non-Bank aid flows.⁷ Moreover, Kilby (2005) considers only one aspect of economic policy, deregulation.⁸

3. Data and Methodology

3.1. Dependent variable and variables of interest

In this study we analyze the association of World Bank lending with the quality of economic policy. In contrast with most of the existing literature on policy lending, our dependent variable is not a final outcome measure but rather the quality of economic man-

agement, as measured by the World Bank’s Country Policy and Institutional Assessment (CPIA) ratings. The CPIA assessments are subjective ratings of 16 policy indicators, grouped into 4 “clusters”, updated annually by World Bank staff.⁹ Possible scores on each indicator range from one to six, including half-point increments (for example 3.5). For this analysis, our main dependent variable is the simple average of CPIA clusters A and B, which broadly reflects the so-called “Washington Consensus” neo-liberal policy prescriptions (Williamson, 1994). Cluster A covers macroeconomic and debt policy, while cluster B addresses structural policies, including trade, financial sector policies, and regulation of private enterprise.¹⁰ The mean score of this CPIA-based policy quality indicator in our sample is 3.61, with a standard deviation of 0.73.

The CPIA is arguably the most appropriate policy measure, because its content reflects the views of World Bank management and staff regarding what policies are most conducive to poverty reduction and the effective use of aid resources. Admittedly, there are prominent skeptics of the development efficacy of neo-liberal policy prescriptions (see, for example, Rodrik, 2006). The CPIA criteria may be seen as representing only one particular view on what constitutes sound economic policy, and the policy prescriptions reflected in these ratings may not necessarily lead to the desired outcomes of growth and poverty reduction. Regardless of any perceived deficiencies in the CPIA’s content, it is the most relevant available cross-country indicator of the policies World Bank country teams are attempting to achieve when they design DPLs.

The CPIA indicators reflect the subjective judgments of World Bank staff. However, they are correlated with conceptually-related objective indicators, as well as with subjective indicators produced by other organizations. The CPIA cluster A and B average is correlated in the expected direction with macroeconomic indicators such as inflation ($r = -0.12$) or government debt ($r = -0.43$). It is also strongly correlated with the International Country Risk Guide’s (ICRG) “economic risk” composite – an index including GDP

per capita, real GDP growth, annual inflation rate, budget balance and current account balance as components (see the corresponding figure in the online appendix).

In robustness tests we supplement the CPIA with alternative measures of neoliberal economic policies from the Fraser Institute and Heritage Foundation.¹¹ Replicating results for these alternative dependent variables is useful for two reasons. First, it shows that the CPIA does not represent a particularly idiosyncratic World Bank view of what good policies look like. On the contrary, there is quite a bit of conceptual overlap with the Fraser and Heritage “economic freedom” indexes. Similarly to the CPIA’s four “clusters”, Fraser’s Economic Freedom of the World (EFW) index groups indicators into five policy “areas”: size of government, secure property rights, access to sound money, freedom to trade internationally, and regulation of credit, labor and business. The Heritage’s Index of Economic Freedom covers ten components which are grouped in four categories: rule of law, limited government, regulatory efficiency and open markets. Again, this categorization closely resembles the subdivisions found in the CPIA. Empirically, there is also a close match. The pairwise correlations for the year 2008 between CPIA and EFW, and CPIA and Heritage, are 0.68 and 0.71 respectively.

A second reason to test our model with alternative dependent variables is to avoid capturing any spurious correlation. Specifically, replicating our main results with the EFW and Heritage indexes rules out the possibility that positive correlations between DPLs and progress on economic policy reform are an artifact of CPIA ratings bias. The CPIA ratings process for a given country involves numerous World Bank staff, potentially including those involved in designing, approving or supervising DPLs to the country. Despite multiple levels of reviews in the CPIA process, it is possible that country teams implementing a DPL will have an over-optimistic view of the loan’s impact, and try to increase subsequent CPIA ratings beyond what is justified by actual results. The Heritage and Fraser indicators are immune to this potential bias. Note that our 2SLS tests, instrumenting for DPLs, will

also correct for this potential bias, even when using CPIA as the dependent variable.

Another related variable is the measure on regulatory quality from the Worldwide Governance Indicators (Kaufmann et al., 2010). It captures the ability of the government to formulate and implement sound policies for private sector development. The correlation with CPIA cluster B in our sample is high, 0.85. This variable also generates significant results in a regression setting.¹² However, we prefer using the economic freedom indexes mentioned above as they conceptually cover CPIA cluster A and B, while WGI's measure only covers part of cluster B, that is, regulation of private enterprise.

Even if real improvements in policy are associated with DPLs, it is possible they would have occurred anyway, even in the absence of the lending program. In the new operational policy (OP 8.60), the basic rationale of a DPL is that the prospect of receiving a loan motivates a government to implement a set of “prior actions” (policy conditions negotiated with the Bank), and funds are then disbursed in anticipation of further reforms. One might argue that improvements in policy (as measured by the CPIA) can result merely from a government implementing a set of prior actions that were already planned or underway before any discussion of a DPL began. However, prior actions tend to include “de jure” reforms – such as passing a law or creating a new office – that would rarely be significant enough to warrant an increase in a CPIA rating. Prior actions are usually designed to represent a signal of commitment, or “first installment” in a larger package of reforms supported by a DPL. The majority of completed DPLs are rated by the Bank’s Independent Evaluation Group (IEG) as being successful in attaining their objectives, and a loan that accomplishes nothing more than the implementation of its prior actions does not necessarily receive a favorable rating.¹³ Our 2SLS and GMM tests correct for the possibility that countries receiving DPLs might tend to be the same ones that would have reformed most successfully even in the absence of a loan.

Following Easterly (2005), our key variable of interest is the cumulative number of

policy loans. That is, we focus on repeated lending to the same country, since we believe supporting policy change is a multistage and long term process. However, unlike Easterly (2005), who included all development policy loans in his analyses of macroeconomic policy distortions, we consider only the subset of loans that support policy reforms in the areas measured by CPIA clusters A and B. These loans – which henceforth we will call “market reform loans” – comprise less than sixty percent of the Bank’s total development policy lending portfolio (see the corresponding table in the online appendix). Market reform loans are not evenly distributed across countries: Ghana tops the list with a total of 17 loans (see the corresponding figure in the online appendix). Among the countries that have received at least one market reform loan, the median number of cumulative loans is four.

As an alternative to the cumulative number of DPLs, we also consider the number of cumulative loan conditions (or “prior actions”).¹⁴ Again, we count only the conditions related to the content of CPIA clusters A and B. Argentina is clearly an outlying observation (see the corresponding figure in the online appendix), with a total of 336 market reform conditions, mostly from the World Bank’s involvement in Argentina’s large-scale economic reforms during the 1990s and early 2000s (see, for example, Bambaci et al., 2002). We test the effect of conditions on policy reform both with and without this outlier in the sample.

3.2. *Model specifications*

Econometrically, we estimate the following equation:

$$y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 Z_{i,t} + \delta_i + \epsilon_{i,t} \quad (1)$$

where y_{it} is the average of CPIA cluster A and B for country i in year t . X_{it} represents the cumulative number of market reform loans (or conditions) for country i in year t . For both variables, we estimate a linear effect, but also specified a model with diminishing returns as well as a quadratic relation.¹⁵ Z_{it} is a vector of control variables. Aid from other

donors could have direct or indirect effects on policy reform, so we include total aid over GDP as a control variable. Following Besley and Persson (2011) among other studies, we include a measure of democracy, specifically the Freedom House index of political freedoms.¹⁶ We include a time trend, to control for any secular improvements in economic policy independent of any impact of World Bank loans, and for any potential tendency for inflation over time in CPIA ratings. To correct for the possibility that policy quality may be inferred in part from performance, we control for the logarithm of GDP per capita. δ_i are country fixed effects. Descriptive statistics for these variables are presented in the online appendix. We estimate the coefficients of this model by employing OLS on a comprehensive country-year panel of aid recipient countries that runs from 1995 to 2008. Standard errors are adjusted for country clustering of observations.

Because number of loans and conditions are continuous variables, we correct for sample selection using instrumental variables techniques as in Easterly (2005) rather than Heckman selection models. We use two alternative methods. First, we estimate equation 1 with system GMM (Arellano and Bover, 1995; Blundell and Bond, 1998) and instrument our variables of interest with their lagged differenced values.¹⁷ The Arellano and Bond (1991) tests indicate the presence of substantial autocorrelation: though we can reject serial correlation in differences at the five percent level from $AR(5)$ onwards, the p-values for $AR(7)$ and $AR(9)$ are respectively 0.059 and 0.089 with the number of cumulative loans as the key independent variable. For the number of conditions variable, the p-value drops below the five percent level for $AR(7)$ to 0.036. Hence, we lag our variables of interest to the highest extent possible, that is, 15 periods. Furthermore, as the number of time periods grows large, the instrument count increases exponentially, making results about estimators and related specification tests invalid (Roodman, 2009). One solution to this problem is to use only certain lags. Thus, we limit the number of lags per time period to one. In order to minimize correlation across countries in the idiosyncratic errors, we also include time

dummies instead of a time trend.¹⁸

The validity of this identification strategy rests on the assumption that $E(\Delta x_{i,t-j}(\delta_i + \epsilon_{i,t})) = 0$, which implies that changes in the receipt of policy loans are uncorrelated with the fixed effects. According to the Diff-in-Hansen tests reported in table 6, these additional moment conditions are met. Next, system GMM estimators may suffer from weak instrument biases (Bun and Windmeijer, 2010). Unfortunately, there are no formal tests available to evaluate instrument strength. Hence, we assume that the internal instruments used here are sufficiently strong to identify the effect of policy lending on the quality of policies.

Given the restricted assumptions required for system GMM, we use a cross-sectional version of the dataset and employ 2SLS as a second correction for possible selection bias.¹⁹ We estimate the following cross-sectional equation:

$$\Delta y_{i,t} = \gamma_0 + \gamma_1 y_{i,t_0} + \gamma_2 \Delta \hat{X}_{i,t} + \gamma_3 Z_{i..} + v_i \quad (2)$$

The dependent variable here is the change in policy quality, measured over the period 1996-2008.²⁰ A convenient implication of using the change in policy quality as the dependent variable is that time-invariant heterogeneity between countries – for example, colonial heritage, legal tradition, and cultural norms – should matter relatively little, as most of its effects will arguably be captured by the initial level of policy quality.

Key independent variables are the logarithm of the number of cumulative market reform loans (or conditions) from 1996 through 2008. In the first stage we instrument for number of loans or conditions with the logarithm of population in 1996 and the average fraction of key votes in the UN General Assembly (UNGA) aligned with the G-7 over the period 1995-2008. Population size is used as an instrument for aid in many other researches (for example, Boone, 1996; Burnside and Dollar, 2000; Djankov et al., 2008) and has

been shown to be statistically unrelated with a wide variety of institutional indicators (Rose, 2006). However, its use is not without problems, especially when economic growth is the dependent variable. Most notably, Bazzi and Clemens (2013) argue that country size may affect growth through multiple (endogenous) channels – trade, FDI, aid, ... – which turns population as an instrument invalid when those factors are omitted or not properly instrumented for. The fact that many empirical studies have used population as an instrument for some endogenous variable without including other channels confirms their point. Yet, where the *quality of policies* – that is, how policies are designed and not the actual policy outcomes – is concerned, their conjecture is less obvious and without much empirical support. Obviously, it is possible that orthogonality conditions are not met, even though test statistics for overidentification are reassuring. As a second instrument we use the average fraction of key votes in the UN General Assembly (UNGA) aligned with the G-7. Here we follow, among others, Barro and Lee (2005) and Kilby (2013), who draw on the literature that examines the influence of powerful nations on IFI decision making. This line of research has convincingly shown that important donors use IFI resources to seek support of strategically relevant countries. Again, for the instrument to be valid, we assume that UN voting decisions are not directly related to the economic policies countries opt for.

As controls we include the initial level of policy quality, average annual aid as a share of GDP, and average annual growth in GDP per capita over the period 1996-2008, the logarithm of initial income per capita, a measure for ethnic fractionalization (Alesina et al., 1999; Collier, 2000), initial political freedom and the change in political freedom over the period 1996-2008. See the online appendix for descriptive statistics. The coefficients of equation 2 are estimated using 126 observations, one for each country for which CPIA data are available from both 1996 and 2008. In the next section we discuss our empirical findings.

4. Empirical Findings

4.1. Baseline results and spillovers

Table 1 presents the results for the number of market reform loans. Number of loans is significantly related to policy quality in each of the three specifications – linear, quadratic and logarithmic. In the linear specification (table 1, equation 1), each additional market reform loan is estimated to increase the CPIA score by .07 on average. Results for the quadratic model imply that the maximum improvement in CPIA (relative to the case of no DPLs) is about 0.90, corresponding to the case of 13 loans. For the logarithmic specification, a first loan increases the CPIA score by 0.40 points on average, and a second loan by 0.21 points. However, the reported goodness-of-fit measures suggest that the logarithmic specification is most appropriate. Furthermore, both the J-test and Cox-Pesaran test for non-nested models indicate that the model with positive but diminishing returns to more DPLs better fits the data than the linear and quadratic models.²¹ The graphical output of a semiparametric estimation – see figure 2 – further confirms the choice of the logarithmic model. For space considerations, we will therefore report only the findings of the logarithmic model in subsequent regressions when the number of loans is the main variable of interest. Table 1 also reports a significant negative time trend over the 1995 to 2008 period. Higher per capita income and higher aid/GDP are associated with better economic policies. Political freedoms are not significant, perhaps in part due to limited variation in the data over time for many countries, coupled with the inclusion of country fixed effects.

Table 2 reports findings for the number of cumulative conditions. The first equation presents the results from estimating the quadratic specification using the full sample. A highly significant concave relation appears, with a predicted turning point at 149 cumulative conditions - equal to three times the average number in our sample, and two standard deviations above the mean. However, figure 1 – the partial residual plot for the number of

cumulative conditions – suggests that Argentina is a highly influential case in estimating this relationship. Without Argentina in the sample (table 2, equation 3), the coefficient on number of conditions squared declines and is no longer significant at conventional levels. The estimated turning point drops from 149 to 20 cumulative conditions. Because Argentina is an extreme outlying and influential observation (in the quadratic specification), we drop it from the sample in subsequent tests.

Equations 2 and 4 of table 2 show that the coefficient for the number of cumulative conditions is positive and significant in both the linear and logarithmic specifications. According to equation 2, one additional market reform condition increases the CPIA score for the typical country by 0.04 points. The logarithmic model predicts that the first market reform condition increases the CPIA score by 0.11 points on average. Table 2 also shows that control variables behave in similar fashion as in table 1: income and aid are positively associated with policy quality, and controlling for other variables there is a significant negative time trend. Concerning model fit, neither the reported goodness-of-fit measures, nor the J or Cox-Pesaran test, nor the semiparametric estimation (see figure 3) provide robust indications which specification has the best fit. For the number of cumulative conditions, we thus report all three specifications for most tests.

Next, we also check whether the implementation of market reform programs has “crowded out” policy improvements in other areas. We do so by substituting the CPIA social policy (CPIA C) and public sector governance (CPIA D) cluster averages for clusters A and B as dependent variables. A priori, there are reasons to expect negative spillovers on other policy areas. For instance, improving policy in one sector might divert rent-seeking activities to other sectors. Also, focusing on one policy area could attract human capital and other resources from other sectors, reducing the ability to design and implement adequate policies in those sectors. On the other hand, new rules and norms of behavior in one part of the public sector might transplant to other departments or agencies (see, for example,

Banerjee, 1992; Mullainathan, 2006). Thus we might also expect some “crowding in” of reforms, that is, positive spillovers. Table 3 however shows that neither loans nor conditions designed to improve policies related to clusters A and B have any significant net impact on CPIA cluster C or cluster D. Coefficient signs in the CPIA C regressions are consistent with positive spillovers, but p-values are above conventional significance levels. Coefficient signs are mixed in the CPIA D regressions, and none come close to significance. One possible explanation for the lack of (positive) spillovers is the length of the governance results chain. That is, while improvements in cluster A and B are often characterized by a short chain from inputs to outputs - for example, “stroke-of-the-pen” reforms such as reduction in trade tariffs - the governance results chain in other areas, such as tackling corruption, is much longer and thus harder to influence (World Bank, 2013).

4.2. Sample restrictions, additional controls and alternative dependent variables

In this subsection, we conduct several robustness checks. First, we employ two sample restrictions. We follow Easterly (2005) in limiting the sample to include only countries that have received at least one economic policy loan over the period 1980-2010. With this change, about one sixth of observations (and countries) are dropped. Selection bias should be reduced – but not eliminated entirely – in this more homogeneous sample. As equation 1 of table 4 shows, the coefficient on (the log of) the cumulative number of loans remains positive and highly significant, although it is somewhat smaller in magnitude than in table 1, equation 3. As shown in the first row of table 5, the number of conditions remains significant only in the linear specification.

As an alternative sample restriction, we drop all observations for a country after the last market reform loan to that country has closed.²² About one third of observations are dropped with this change. If reforms associated with DPLs are often not sustained following completion of the loan, then the estimated effects should increase when the years following loan closing are dropped. Equation 2 of table 4 indicates that the impact

of economic policy lending is slightly higher, with similar significance levels, with this restriction (.444, compared to .406). Coefficients are also slightly larger for the number of cumulative conditions, as shown in the second row of table 5. Although the coefficients decline in magnitude only slightly with this sample restriction, these patterns are consistent with the conjecture that there is some backsliding of reforms after the loans are fully disbursed.

Next, we test whether results are robust to including additional controls. Chauvet and Collier (2009) find that elections matter for economic policy and distinguish between the frequency effect of elections and the cyclical effect of elections. Dreher et al. (2009) hypothesize that debt incurred in the run up to elections will increase the likelihood of a World Bank loan, but show empirically that World Bank loans are actually less frequent in the wake of an election. We therefore add a measure of elections frequency, a variable that captures the stage of the political business cycle, and a dummy for lagged elections. We include debt service, as it is found to affect the number of World Bank projects a country receives (Dreher et al., 2009). We also add a dummy variable coded 1 if a country signed an agreement with the IMF. Finally, we include (the log of) population, with no theoretical prior but simply to control for possible economies or dis-economies of scale in policy reform. Inclusion of these variables may correct for any omitted variable bias. We also control for gross IDA disbursements, as a correction for one potential source of reverse causation. Countries with higher CPIA ratings receive higher allocations of IDA aid, other things equal, which in turn may increase the likelihood of receiving a DPL. Because any causal effect of CPIA ratings on DPLs is mediated by IDA disbursements, controlling for the latter will effectively correct for this potential source of endogeneity bias. In the next subsection we will treat endogeneity concerns in a more general way. Equation 3 of table 4 shows that the (log of the) number of cumulative market reform loans remains positively and significantly related to the quality of economic policy. The coefficient magnitude (.331)

is reduced somewhat, but it is not directly comparable to equation 3 of table 1, because missing data on some of the additional control variables reduces the sample by nearly one third. Among the added control variables, only IDA volumes are significant: as expected, they are positively related to CPIA ratings. As shown in the third row of table 5, with these additional controls the coefficient for the number of cumulative conditions remains positive and highly significant in the linear specification.

As CPIA ratings are produced within the World Bank, one might argue that results could be driven by spurious correlation, for example if CPIA scores for a country are inflated to justify more lending in general, and/or to justify providing loans in the form of budget support. For this reason, we show that our main results are robust to using two alternative dependent variables, from the “economic freedom” indexes developed by the Fraser Institute and the Heritage Foundation. For both variables, we aggregate certain subindices to correspond as closely as possible to the questions in CPIA clusters A and B. Equations 4 and 5 of table 4 show that we again find a significantly positive effect of World Bank lending on the quality of economic policy.²³ Number of conditions has a positive and significant coefficient in both the linear and logarithmic specifications for the Fraser Institute index, as shown in the fourth row of table 5. For the Heritage Foundation index (last row of table 5), the quadratic specification provides the best fit between number of conditions and quality of economic policy. The maximum increase in the Heritage index (by 8 points, or nearly one standard deviation) is estimated to occur at 127 conditions. Beyond 254 conditions, policy lending becomes detrimental, relative to the case of no conditions at all. In our data set only 17 out of the 117 countries that received at least one market reform condition lie beyond the predicted turning point. World Bank conditionality was detrimental for only one country (Argentina), according to this specification.

4.3. Endogeneity of policy lending

In this subsection we provide a more general correction for endogeneity of policy lending in two different ways. First, we correct for endogeneity by employing system GMM in the panel dataset. Because the Arellano and Bond (1991) tests indicate the presence of substantial autocorrelation, we lag our variables of interest to the highest extent possible, that is, 15 periods. Furthermore, in order to limit the total number of instruments, we select a lag range of one. Results are presented in table 6. For comparability, we only report the findings of the logarithmic model.²⁴ Coefficients are positive and significant for both the number of loans (equation 1) and the number of conditions (equation 2). Furthermore, test statistics presented at the bottom of table 6 are reassuring. The p-values of the Hansen J statistic do not reject the null that instruments are exogenous. The values reported for the Diff-in-Hansen test provide an indication whether the additional moment restrictions necessary for system GMM are met (Bond et al., 2001). With p-values of around 0.45 for both variables, we do not reject the null that the additional moment conditions are valid.

As a second robustness test, we employ 2SLS and estimate equation 2 in a cross-sectional version of the data. With the panel dataset, we are limited to using mechanical instruments in GMM, because substantive instruments that significantly predict DPLs exhibit little or no time series variation. Moving to cross section data allows us to avoid that problem. The dependent variable here is the change in CPIA cluster A and B, and the endogenous regressor is the logarithm of the number of cumulative loans (or conditions), both measured over 1996 to 2008. In the first stage we instrument for number of DPLs (or conditions) with (the log of) population (in 1996) and the average fraction of the country's key votes in the UNGA that are aligned with the votes of G-7 countries over the period 1995-2008 (Barro and Lee, 2005; Kilby, 2011). We expect larger countries, and allies of major donors, to receive more DPLs. We assume neither variable directly affects quality of economic policies; note population was not significant when added as a control variable

to equation 3 of table 4.

Results for OLS and 2SLS regressions are reported in tables 7 and 8. Equation 1, table 7 shows that the effect of loans on changes in policy quality is positive and statistically significant. Furthermore, the coefficient for initial level of policy quality is significantly negative, implying a regression toward the mean effect. Both the initial level of political rights and its change over the period are associated with improved policy quality²⁵. This finding is consistent with Svensson (2003) and Heckelman and Knack (2008), but inconsistent with other studies suggesting that democratic institutions might actually hamper reform (see, for example, Alesina and Drazen, 1991; Rodrik, 1996). Equations 2 and 3 present the results from 2SLS estimation. Equation 2 shows first-stage results. Population and UN voting are both highly significant predictors of more loans. The F-statistic of excluded instruments is 19.12, which indicates a strong association of our instruments with the receipt of World Bank DPLs. Furthermore, Wooldridge (1995)’s robust score test of overidentifying restrictions does not reject the null that the excluded instruments are exogenous to the quality of policy (test score = 0.21, p-value = 0.64). In equation 3, the exogenous effect of policy lending is reported. The coefficient on loans more than triples in comparison with its OLS counterpart, suggesting that the net effect of endogeneity bias was negative. The 2SLS regression confirms the regression toward the mean effect. In addition, both the initial income level and income growth now have a positive and significant effect on changes in policy.

Table 8 presents the 2SLS results when the number of cumulative conditions is substituted for number of loans as the key regressor. Again, regression diagnostics support our identification strategy. The first-stage F-statistic is 27, and the p-value for the overidentification test is .906. As table 8 shows, findings are similar to results in table 7. The OLS coefficient on log of conditions is positive and highly significant (equation 1), but it nearly triples in magnitude when we instrument for conditions with initial population and UNGA

voting. The coefficient on initial CPIA is again negative and statistically significant, implying that, on average, countries with greater initial policy quality tend to improve less over time. Furthermore, estimates suggest that increasing political rights improves economic policy. The 2SLS regression also confirms that economic policy improvements are associated with high initial income and income growth.

5. Summary and Concluding Remarks

In this study we investigate the impact of World Bank policy loans on the quality of economic policy, correcting for several methodological problems and allowing for the possibility of increasing or decreasing returns to additional loans or conditions. We find that policy lending has a positive but diminishing effect on the quality of economic policy. Results are robust to sample restrictions, additional controls, the use of alternative indicators of the quality of economic policy, and correction for endogeneity with system GMM and cross-sectional 2SLS. Similar results are generally obtained when we substitute the number of cumulative conditions for the number of cumulative loans, although in this case no one functional form consistently best fits the data. There is some evidence for negative returns to additional conditions beyond some point, but the estimated inflection point is highly sensitive to the inclusion or exclusion of Argentina in the sample. The average number of conditions in DPLs declined from about 35 in the 1980s to about 12 by 2005, and our results provide some support for the Bank's decision to make conditionality less onerous. Finally, we investigate the possibility of spillover effects on other policy areas, and show that investing in economic policy reform does not significantly affect policy quality for good or ill in the areas of public sector governance, and human development, social policy, and environmental policy.

Our main results are in contrast with most of the research examining the effectiveness of adjustment lending. Although there are many differences in data and methodology that

could explain this discrepancy, four of them are particularly worthy of note. First, estimating the impact of development policy lending calls for a sound identification strategy. However, many of the early studies employed a before-after analysis or a with-without approach using strong but dubious assumptions. In contrast, our study relied on instrumental variables techniques to obtain identification. Second, our analysis distinguished among the policy targets of DPLs – many of them target sectoral policies, not economic policies. Failing to make this distinction can produce a downward bias in the estimated impact of lending on policy reform. In this respect, our study is similar in spirit to Clemens et al. (2012), who show that aid’s estimated impact on short-run growth strengthens when humanitarian and other components of aid are excluded that are not intended to further short-run growth. Third, instead of looking at final outcome measures such as economic growth – for which aid might only represent a weak signal (Roodman, 2007) – we take as the dependent variable what World Bank country teams are attempting to achieve when they design DPLs, that is, the quality of development policies. And finally, the time period under investigation is different. Most research evaluates the first two decades of adjustment lending. However, as mentioned in section 2 the practice of development policy lending evolved substantially over time, particularly since the end of the 1990s.

The more positive results in our study suggest that the World Bank’s claims about learning from its mistakes with traditional adjustment lending have some validity. Furthermore, even though we do not look at economic growth directly, there is good reason to believe that the policy reforms the World Bank supports stimulate development.²⁶ Consequently, finding that the Bank’s key lending instrument for improving economic policies appears to be effective is important for its twin goals of ending extreme poverty and stimulating shared prosperity.

Finally, our analysis considered only policy lending conducted by the World Bank. We would welcome similar studies of aid programs intended to promote policy reform, by the

regional development banks, the European Union or large bilateral donors, using their own internal datasets where available.

Acknowledgement

We would like to thank Vincenzo Verardi, Adam Wagstaff, Peter Moll, Patricia Geli, the seminar participants at the 2013 LAGV conference and one anonymous referee for useful comments and suggestions. Lodewijk is also indebted to the Institute of Development Policy and Management (IOB) and the Research Foundation Flanders (FWO) for financial support. When requested, we are happy to provide all publicly available data and programming code for replication purposes.

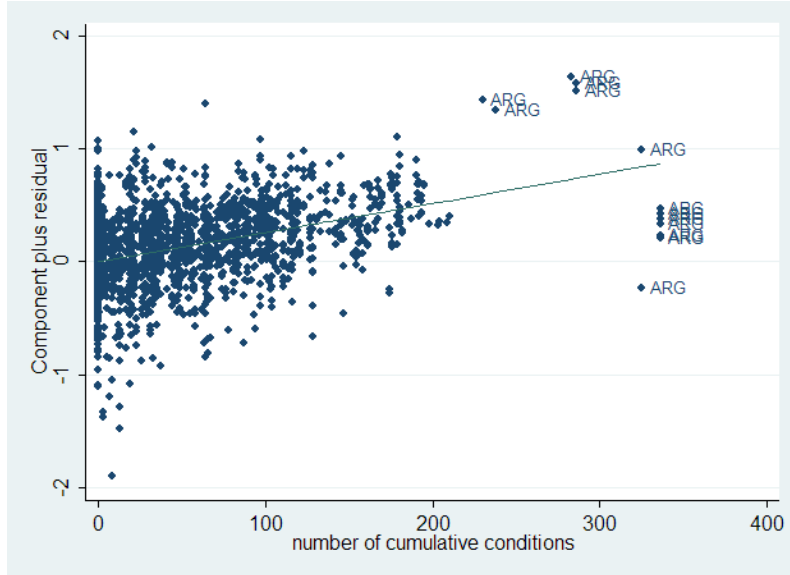
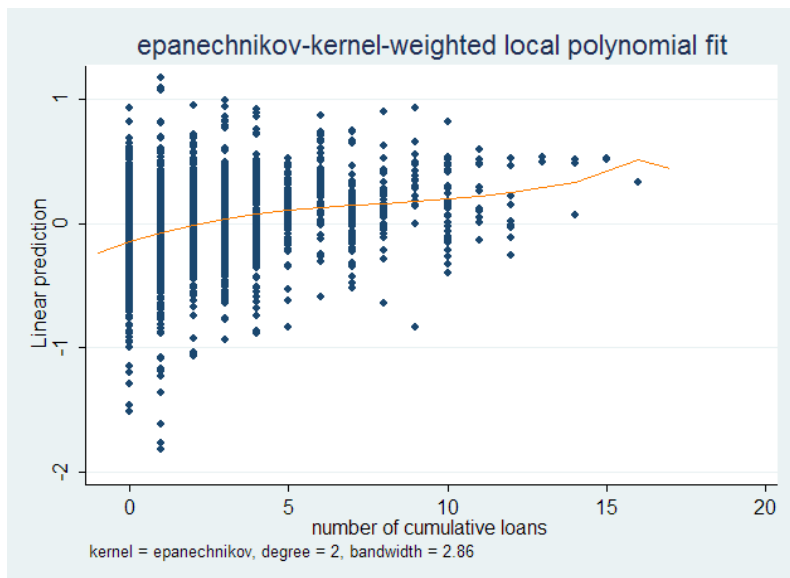


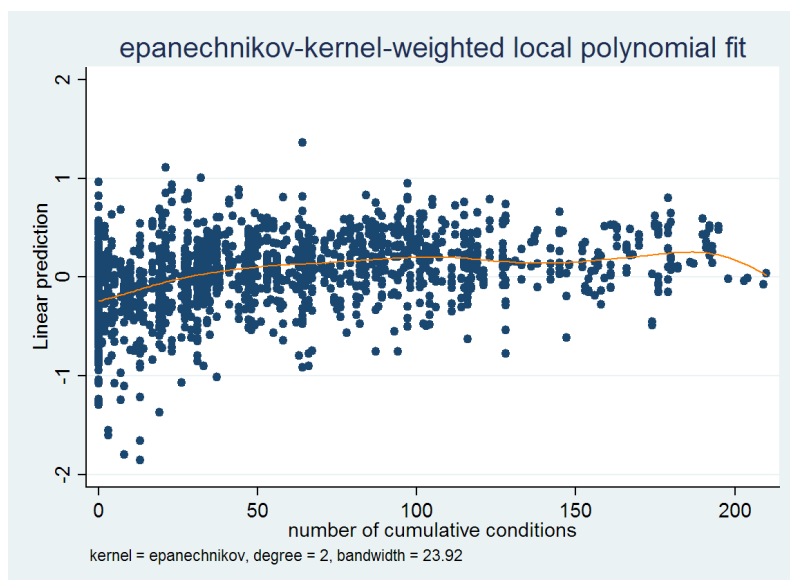
Figure 1: partial residual plot of number of cumulative conditions based on equation 1, table 2, with Argentina included

Figure 2: Non-parametric fit of cumulative loans



Note: semiparametric fixed-effects regression using STATA's xtsemipar command with CPIA cluster A and B average as dependent variable, log of per capita GDP, aid over GDP, political rights and a time trend as parameterized variables and cumulative loans as non parameterized variable. Polynomial of degree two fitted. Standard errors clustered by country.

Figure 3: Non-parametric fit of cumulative conditions



Note: semiparametric fixed-effects regression using STATA's `xtsemipar` command with CPIA cluster A and B average as dependent variable, log of per capita GDP, aid over GDP, political rights and a time trend as parameterized variables and cumulative conditions as non parameterized variable. Polynomial of degree two fitted. Standard errors clustered by country. Argentina excluded from the sample.

Table 1: panel regression of CPIA clusters A and B average on cumulative loans

equation no.	(1)	(2)	(3)
number of cumulative loans	.073 (.023)***	.134 (.047)***	.
number of cumulative loans (squared)	.	-.005 (.003)*	.
log of number of cumulative loans	.	.	.406 (.112)***
year	-.023 (.008)***	-.024 (.008)***	-.026 (.008)***
log GDP per capita (PPP)	.805 (.152)***	.799 (.149)***	.814 (.148)***
aid over GDP	1.618 (.531)***	1.577 (.519)***	1.512 (.515)***
Political Rights	-.016 (.021)	-.015 (.021)	-.011 (.021)
<i>country fixed effects</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Observations	1761	1761	1761
Countries	139	139	139
R^2	.134	.139	.147
Adjusted R^2	.131	.137	.144
AIC	1113.115	1103.17	1086.232
BIC	1140.483	1136.012	1113.601

Note: * significance at 10%; ** significance at 5%; *** significance at 1%.

Table 2: panel regression of CPIA clusters A and B average on cumulative conditions

equation no.	(1)	(2)	(3)	(4)
variation	quad.+Arg.	linear	quad	log
number of cumulative conditions	.010 (.003)***	.004 (.001)***	.007 (.003)**	.
number of cumulative conditions (squared)	-.00003 (.00001)***	.	-.00002 (.00001)	.
log of number of cumulative conditions111 (.061)*
year	-.018 (.008)**	-.017 (.008)**	-.018 (.008)**	-.016 (.008)**
log GDP per capita (PPP)	.807 (.155)***	.792 (.154)***	.794 (.154)***	.816 (.158)***
aid over GDP	1.697 (.520)***	1.682 (.535)***	1.690 (.526)***	1.689 (.521)***
Political Rights	-.016 (.021)	-.019 (.021)	-.016 (.021)	-.016 (.021)
<i>country fixed effects</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Observations	1761	1748	1748	1748
Countries	139	138	138	138
R^2	.14	.124	.127	.123
Adjusted R^2	.137	.122	.124	.12
AIC	1102.59	1090.29	1085.58	1092.87
BIC	1135.43	1117.46	1118.38	1120.12

Note: * significance at 10%; ** significance at 5%; *** significance at 1%. Argentina excluded from the sample for equations (2) through (4).

Table 3: spillover effects on other policy areas

dependent variable	CPIA C	CPIA D
log of number of cumulative loans	.153 (.114)	.103 (.088)
cumulative conditions, linear spec.	.0005 (.001)	-.0003 (.001)
cumulative conditions, quadratic spec.	.001 (.004)	-.0004 (.004)
	-.000003 (.00002)	.0000005 (.00002)
cumulative conditions, logarithmic spec.	.037 (.058)	-.024 (.052)

Note: Regression results from estimating equation 1 with CPIA C and CPIA D as dependent variables. CPIA cluster C average measures the quality of policies for social inclusion and equity and CPIA cluster D average measures the quality of policies for public sector governance. Only coefficient estimates and clustered standard errors of loans and conditions variables reported. Argentina excluded from the sample when the number of conditions is used as variable of interest. When the number of loans is used as variable of interest, 1761 observations are used covering 139 countries. When the number of conditions is used as variable of interest, 1748 observations are used covering 138 countries.

Table 4: robustness tests: cumulative loans

variation	DPL>0	closing year	controls	EFW	Heritage
equation no.	(1)	(2)	(3)	(4)	(5)
log of number of cumulative loans	.341 (.115)***	.444 (.127)***	.331 (.129)**	.394 (.134)***	2.623 (1.413)*
year	-.018 (.009)**	-.039 (.011)***	-.036 (.019)*	.045 (.009)***	-.130 (.101)
log GDP per capita (PPP)	.727 (.148)***	.980 (.260)***	.899 (.263)***	.442 (.204)**	12.193 (2.566)***
aid over GDP	1.564 (.556)***	1.458 (.576)**	2.019 (.706)***	1.961 (.973)**	-34.705 (11.410)***
Political Rights	-.008 (.021)	-.012 (.026)	.010 (.021)	-.079 (.042)*	-.751 (.358)**
log of gross IDA	.	.	.077 (.023)***	.	.
IMF arrangement	.	.	-.022 (.025)	.	.
debt service (% of GNI)	.	.	-.004 (.004)	.	.
log of population	.	.	.941 (.734)	.	.
lagged election	.	.	-.009 (.019)	.	.
election cycle	.	.	-.011 (.014)	.	.
election frequency	.	.	.008 (.010)	.	.
<i>country fixed effects</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Observations	1443	1182	1235	1040	1607
Countries	114	121	105	95	132
R^2	.143	.136	.191	.547	.195

Note: Panel regression results from several robustness tests. Dependent variable: CPIA clusters A and B average. Variable of interest: (log of) number of cumulative market reform loans. * significance at 10%; ** significance at 5%; *** significance at 1%. Standard errors clustered by country.

Table 5: robustness tests: cumulative conditions

robustness test	linear	quadratic	logarithmic
DPL>0	.002 (.001)*	.	.
	.	.006 (.003)*	.
	.	-.00002 (.00002)	.076 (.060)
closing year	.004 (.001)***	.	.
	.	.008 (.003)**	.
	.	-.00002 (.00002)	.119 (.068)*
controls	.005 (.002)***	.	.
	.	.009 (.004)***	.
	.	-.00002 (.00002)	.126 (.078)
EFW	.007 (.002)***	.	.
	.	.008 (.004)*	.
	.	-.000006 (.00002)	.096 (.057)*
Heritage	.031 (.020)	.	.
	.	.127 (.049)***	.
	.	-.0005 (.0002)**	.757 (.766)

Note: Panel regression results from several robustness tests. Dependent variable: CPIA clusters A and B average. Variable of interest: number of cumulative market reform conditions. Only coefficient estimates and clustered standard errors of conditions variable are reported. * significance at 10%; ** significance at 5%; *** significance at 1%. Argentina excluded from the sample. Taking into account the exclusion of Argentina, observations and number of countries are similar to table 4.

Table 6: System GMM

equation no.	(1)	(2)
log of number of cumulative loans	.194 (.091)**	.
log of number of cumulative conditions	.	.090 (.043)**
log GDP per capita (PPP)	.221 (.063)***	.219 (.062)***
aid over GDP	-1.089 (.528)**	-.947 (.572)*
Political Rights	-.104 (.027)***	-.106 (.027)***
<i>country fixed effects</i>	<i>yes</i>	<i>yes</i>
<i>year fixed effects</i>	<i>yes</i>	<i>yes</i>
Observations	1761	1748
Countries	139	138
Number of instruments	44	44
Wald statistic	159.88	153.02
<i>p-value</i>	0.0001	0.0001
Hansen J-test	23.10	25.01
<i>p-value</i>	0.627	0.518
Diff-in-Hansen test	13.99	13.89
<i>p-value</i>	0.45	0.451

Note: Dependent variable: CPIA cluster A and B average. For equation (1), the log of the number of cumulative loans is the variable of interest. For equation (2), the log of the number of cumulative conditions is the variable of interest. Cluster-robust standard errors are reported. Coefficients estimated with forward orthogonal deviations and level equations for IV style instruments. * significance at 10%; ** significance at 5%; *** significance at 1%. Argentina excluded from the sample when the number of conditions is used as variable of interest.

Table 7: cross-sectional 2SLS number of loans

equation no.	(1)	(2)	(3)
	OLS	First stage	Second Stage
log of cumulative loans 1996-2008	.224 (.065)***	.	.770 (.175)***
CPIA 1996	-.586 (.063)***	.061 (.088)	-.636 (.075)***
average annual GDP per capita growth	.016 (.009)*	.0002 (.005)	.018 (.008)**
log of GDP per capita 1996	.076 (.056)	-.210 (.079)***	.256 (.087)***
ethnic fractionalization	.124 (.176)	.013 (.228)	.068 (.205)
Political Rights 1996	-.058 (.029)**	-.056 (.040)	-.011 (.038)
change in Political Rights	-.102 (.033)***	-.071 (.047)	-.043 (.042)
average annual aid over GDP	-1.635 (1.081)	1.840 (1.969)	.386 (1.295)
log of 1996 population	.	.160 (.041)***	.
average fraction of votes with G-7	.	.896 (.358)**	.
No. observations	126	126	126
R^2	.552	.347	.261
F test of excluded instruments	.	19.1276	.
<i>p-value</i>	.	0.00001	.
test of endogeneity	.	.	13.2256
<i>p-value</i>	.	.	0.0003
Overidentification test	.	.	0.2125
<i>p-value</i>	.	.	0.6449

Note: Dependent variable is the change in policy quality over the period 1996-2008, as measured by the CPIA cluster A and B average. Robust standard errors in parentheses. * significance at 10%; ** significance at 5%; *** significance at 1%.

Table 8: cross-sectional 2SLS number of conditions

equation no.	(1)	(2)	(3)
	OLS	First stage	Second Stage
log of cumulative conditions 1996-2008	.097 (.027)***	.	.274 (.057)***
CPIA 1996	-.588 (.062)***	.146 (.220)	-.630 (.068)***
average annual GDP per capita growth	.016 (.009)*	.003 (.012)	.017 (.007)**
log of GDP per capita 1996	.047 (.053)	-.160 (.195)	.129 (.066)**
ethnic fractionalization	.120 (.177)	.086 (.571)	.070 (.200)
Political Rights 1996	-.063 (.029)**	-.048 (.093)	-.035 (.033)
change in Political Rights	-.118 (.033)***	.031 (.112)	-.103 (.037)***
average annual aid over GDP	-1.706 (1.126)	7.420 (5.021)	-.316 (1.289)
log of 1996 population	.	.419 (.096)***	.
average fraction of votes with G-7	.	2.963 (.847)***	.
No. observations	126	126	126
R^2	.56	.331	.368
F test of excluded instruments	.	27.0558	.
<i>p-value</i>	.	0.00001	.
test of endogeneity	.	.	12.8265
<i>p-value</i>	.	.	0.0003
Overidentification test	.	.	.141
<i>p-value</i>	.	.	0.906

Note: Dependent variable is the change in policy quality over the period 1996-2008, as measured by the CPIA cluster A and B average. Robust standard errors in parentheses. * significance at 10%; ** significance at 5%; *** significance at 1%. Argentina excluded from the sample.

Appendices

Appendix A. Descriptive Statistics, variable definitions and sources

Table A.1: Summary statistics panel model

Variable	Mean	Std. Dev.	Min.	Max.
CPIA cluster A and B average	3.613	0.73	1	5.850
CPIA cluster C	3.422	0.700	1	6
CPIA cluster D	3.226	0.717	1	5.5
EFW	6.313	1.026	2.027	8.932
Heritage	60.09	9.591	20.833	88.183
number of cumulative loans	2.86	2.763	0	16
number of cumulative conditions	50.368	49.784	0	210
year	2001.646	3.883	1995	2008
log GDP per capita (PPP)	7.981	1.018	5.076	10.352
aid over GDP	0.042	0.064	-0.019	0.806
Political Rights	3.777	1.989	1	7
log of gross IDA (current million USD)	2.31	2.26	0	8.311
arrangement with IMF	0.15	0.357	0	1
debt service (% of GNI)	5.021	6.28	0.053	138.888
log of population	15.926	1.756	11.515	20.854
election	0.194	0.396	0	1
election cycle	1.841	2.513	0	24
election frequency	5.151	2.202	1	22

Table A.2: Summary statistics cross-sectional model

Variable	Mean	Std. Dev.	Min.	Max.
change in CPIA cluster A and B	0.046	0.661	-1.731	1.819
log of cumulative loans 1996-2008	.8286652	.7132909	0	2.30259
log of cumulative conditions 1996-2008	2.091848	1.694547	0	4.82831
CPIA A and B 1996	3.699	0.856	1	5.231
average annual GDP per capita growth	4.770	8.168	-2.486	82.035
log of GDP per capita 1996	6.834	1.168	4.191	9.031
ethnic fractionalization	0.473	0.252	0	0.930
Political Rights 1996	3.738	2.06	1	7
change in Political Rights	-0.079	1.312	-5	3
average annual aid over GDP	0.035	0.045	0	0.277
log of population 1996	15.588	1.958	10.618	20.92
average fraction of votes with G-7	0.478	0.203	0.113	0.943

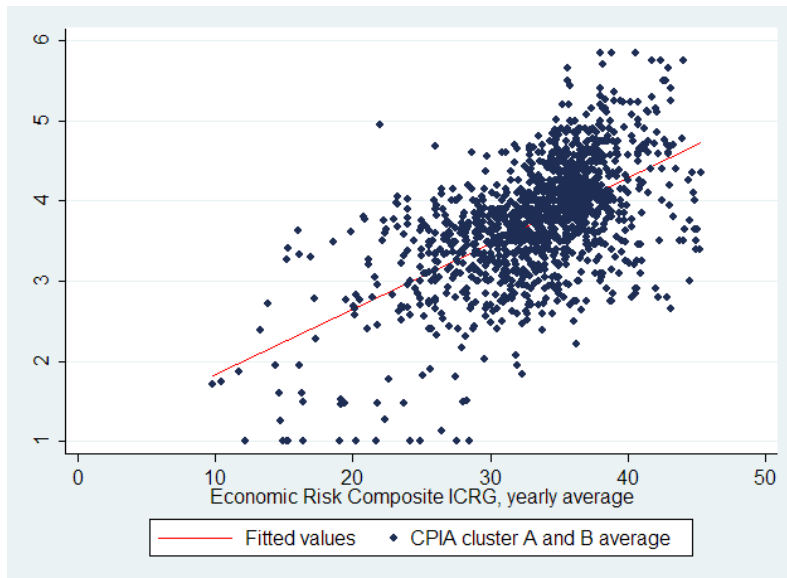


Figure A.4: linear association between CPIA cluster A and B average and ICRG's Economic Risk Composite

Table A.3: sectoral distribution of all effective adjustment loans for the period 1980-2010

	sector	frequency	percentage
Market Reform Loans			
	Economic Policy	450	44.91
Financial and Private Sector Development		121	12.08
	Financial Sector	12	1.2
	Private Sector Development	7	0.7
Other DPLs			
	Agriculture and Rural Development	62	6.19
	Education	29	2.89
	Energy and Mining	46	4.59
	Environment	14	1.4
	Public Financial Management	1	0.1
Global Information/Communications Techn		2	0.2
	Health, Nutrition and Population	8	0.8
	Poverty Reduction	51	5.09
	Public Sector Governance	127	12.67
	Social Development	2	0.2
	Social Protection	49	4.89
	Transport	5	0.5
	Urban Development	14	1.4
	Water	2	0.2
	Total	1,002	100

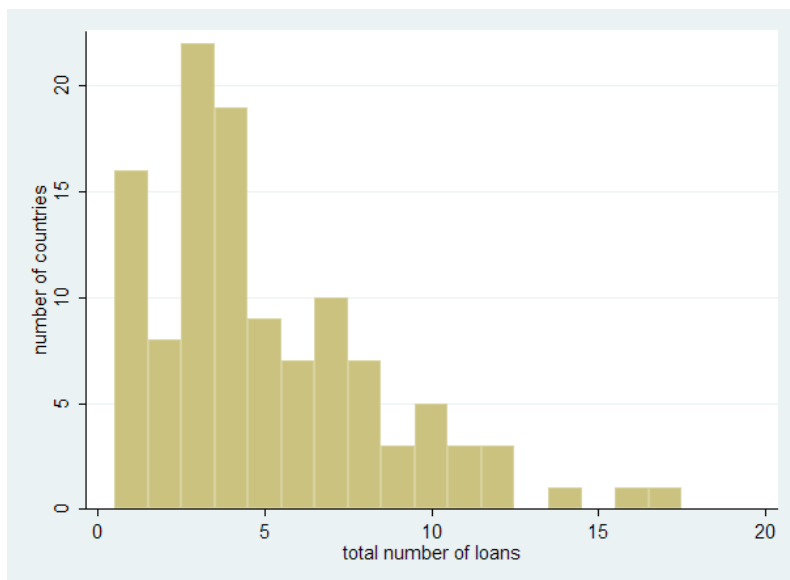


Figure A.5: distribution of cumulative loans for the period 1980-2010

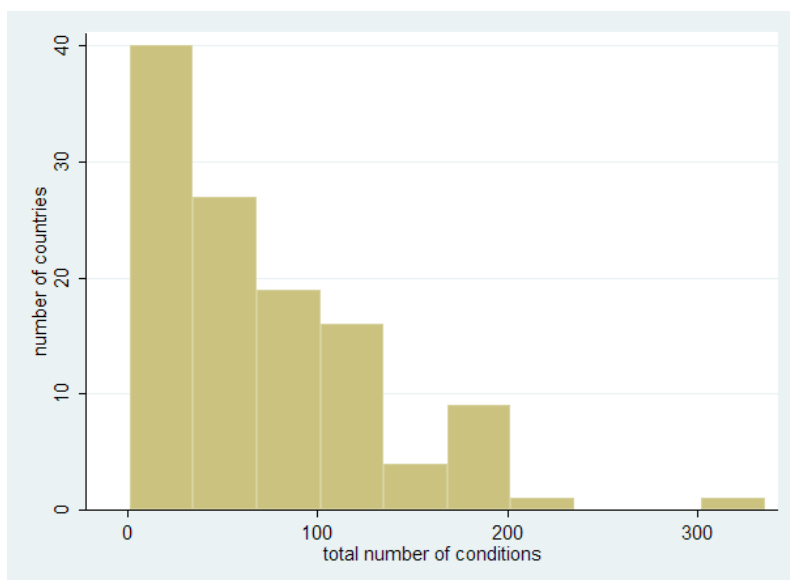


Figure A.6: distribution of cumulative conditions for the period 1980-2010

variable	definition	source
CPIA cluster A	assessment of the quality of a country's economic management	World Bank
CPIA cluster B	assessment of the quality of a country's structural policies	World Bank
CPIA cluster C	assessment of the quality of a country's policies for social inclusion	World Bank
CPIA cluster D	assessment of the quality of a country's public sector management and institutions	World Bank
EFW	average of Economic Freedom of the World Index area 1, 3, 4 and 5	The Fraser Institute
Heritage	Index of Economic Freedom based on trade freedom, business freedom and investment freedom	the Heritage Foundation
number of cumulative loans	cumulative count of loans with sector codes corresponding to CPIA cluster A and B	World Bank
number of cumulative conditions	cumulative counts of conditions with theme codes corresponding to CPIA cluster A and B	World Bank
year	year trend	World Bank
log of GDP per capita	logarithm of GDP per capita, PPP	World Development Indicators
aid over GDP	Net ODA and official aid over GDP (international \$)	based on WDI
Political Rights	measure for Political Rights	Freedomhouse
log of gross IDA	logarithm of one added to gross IDA disbursements, current million USD	based on WDI
arrangement with IMF	dummy coded 1 if an IMF arrangement was signed	Moser and Sturm (2011)
debt service	Total debt service, as % of GNI	WDI
log of population	logarithm of population	WDI
election	dummy coded 1 if an election occurred at t+1	Beck et al. (2001)
election cycle	number of years that separate year t from the nearest election	based on Beck et al. (2001)
election frequency	number of years between election in year t and the previous election	based on Beck et al. (2001)
average annual per capita growth	average annual per capita growth over the period 1996-2008	based on Beck et al. (2001)
ethnic fractionalization	measure for ethnic fractionalization	Alesina et al. (2003)
average fraction of votes with G-7	average fraction of votes on key issues aligned with the G-7 over the period 1995-2008	Dreher and Sturm (2012)

Table A.4: variable definitions and sources

Appendix B. Country Policy and Institutional Assessment

The CPIA scores are designed to measure government policies and institutions, rather than outcomes. The set of criteria are revised periodically to reflect changes in the collective knowledge of practitioners and specialists - both inside and outside the World Bank – regarding policies and public sector management institutions that matter for these outcomes. The criteria are grouped into 4 “clusters” as follows:

- A. Economic Management
 1. Macroeconomic Management
 2. Fiscal Policy
 3. Debt Policy
- B. Structural Policies
 4. Trade
 5. Financial Sector
 6. Business Regulatory Environment
- C. Policies for Social Inclusion/Equity
 7. Gender Equality
 8. Equity of Public Resource Use
 9. Building Human Resources
 10. Social Protection and Labor
 11. Policies and Institutions for Environmental Sustainability
- D. Public Sector Management and Institutions

12. Property Rights and Rule-based Governance
13. Quality of Budgetary and Financial Management
14. Efficiency of Revenue Mobilization
15. Quality of Public Administration
16. Transparency, Accountability, and Corruption in the Public Sector

For each criterion, countries are rated on a scale of 1 (low) to 6 (high). A 1 rating corresponds to a very weak performance, and a 6 rating to a very strong performance. Intermediate scores of 1.5, 2.5, 3.5, 4.5 and 5.5 may also be given. For the years 1995-1997, countries were rated on a scale of 1 to 5. Scores have been rescaled for this research to a scale of 1 to 6. See OPCS (2009) for a detailed elaboration of the scoring procedure.

Appendix C. Additional System GMM regression

Table C.1: Additional System GMM regression

equation no.	(1)	(2)
log of number of cumulative loans	.389 (.10)***	.
log of number of cumulative conditions	.	.146 (.034)***
log GDP per capita (PPP)	.278 (.054)***	.251 (.053)***
aid over GDP	-.485 (.524)	-.481 (.510)
Political Rights	-.089 (.025)***	-.099 (.026)***
<i>country fixed effects</i>	<i>yes</i>	<i>yes</i>
<i>year fixed effects</i>	<i>yes</i>	<i>yes</i>
Observations	1761	1748
Number of instruments	24	24
Wald statistic	194.05	178.86
<i>p-value</i>	0.0001	0.0001
Hansen J-test	7.31	5.77
<i>p-value</i>	0.293	0.449
Diff-in-Hansen test	0.66	2.73
<i>p-value</i>	0.417	0.10

Note: cluster-robust standard errors are reported. Coefficients estimated with forward orthogonal deviations and level equations for IV style instruments. Collapsed instrument matrix. Lags 5 to 10 used for loans and 10 to 15 for conditions. * significance at 10%; ** significance at 5%; *** significance at 1%. Argentina excluded from the sample when the number of conditions is used as variable of interest.

Notes

¹Ex-ante refers to the timing of disbursing conditional loan tranches. With ex-ante disbursement, loan tranches are disbursed before conditions are met, while ex-post disbursement refers to disbursing funds only after prior actions are met.

²World Bank (1989) lists five reasons of why early adjustment lending was so heavily criticized: i) inadequate program design with limited focus on poverty reduction; ii) limited program implementation; iii) programs based on unrealistic assumptions; iv) the weight of SAL on the Bank's lending portfolio; and v) lack of diplomacy and coordination among creditors.

³See Smets et al. (2013) for a recent quantitative analysis concerning the importance of design quality on reform success.

⁴Joseph Stiglitz's address at UNCTAD in 1998 – when he was the Bank's Chief Economist – nicely illustrates the shift in momentum. Consider, for example, the following quote: 'The key ingredients in a successful development strategy are ownership and participation. We have seen again and again that ownership is essential for successful transformation: policies that are imposed from outside may be grudgingly accepted on a superficial basis, but will rarely be implemented as intended [...]. Furthermore, a country's own development strategy provides, then, the overall framework for thinking about a country's plan for change' (Stiglitz, 1998).

⁵This policy shift was formalized in 2004 in a new operational policy, OP 8.60, including the name change from structural adjustment lending to development policy lending. Furthermore, in 2005 the Bank's Development Committee endorsed five good practice principles of policy based lending: country ownership, harmonization with other donors, customization of lending design, criticality of loan conditions, and transparency and predictability of performance. All new development policy operations should adhere to these best practice principles.

⁶Jones et al. (2011) – examining the Bank's support in bringing down tariffs in Eastern Africa – lies somewhere in between as they investigate the period 1992-2002.

⁷Using a set of instruments, Kilby (2005) fails to reject exogeneity of World Bank lending. However, the author points to the limitations of the available instruments and mentions that results should be treated with caution.

⁸Of the five "economic freedom" sub-indexes, he uses only the one on government regulation of business, credit and labor markets.

⁹See OPCS (2009) for a detailed description of the 16 indicators and the assessment procedure used to

generate them.

¹⁰The CPIA overall goes well beyond the Washington Consensus, as cluster C address human development and social and environmental policies, and cluster D covers public sector governance and institutions.

¹¹See Gwartney et al. (2013) and Miller et al. (2013) for a detailed description of both indices. To provide an even closer match with CPIA cluster A and B, we have dropped security of property rights from the Fraser Institute’s index. For the Heritage score, we only retained the following components: openness to trade, government spending, monetary policy, business freedom, investment freedom and financial freedom.

¹²Regression output is available upon request but left out due to space considerations.

¹³As an additional test we dropped from the sample all DPLs that were rated moderately unsatisfactory, unsatisfactory or highly unsatisfactory. As expected, results from regressing the base model on this data turn out more favorably, but are not included due to space considerations.

¹⁴Prior actions are the critical policy conditions that the borrowing government agrees to take for loan tranches to be released. Arguably, some loan conditions may have a larger impact on policy quality than others. Disaggregating conditions by type is beyond the scope of this study, but is an interesting issue for future research.

¹⁵In order to retain the zero observations when making the log transformation, we added 1 to the number of cumulative EP loans and to the number of cumulative prior actions. Results are not sensitive to the specific values added for the log transformations.

¹⁶A referee suggested also to include Freedom House’s civil liberties as a control variable. Due to the high correlation between political and civil liberties, 0.90 in our sample, results changed very little when doing so. Output from these regressions is available upon request.

¹⁷System GMM is mainly used to estimate a dynamic panel model with a lagged dependent variable on the right-hand side. However, it can also be used – as here – to lag endogenous regressors (Roodman, 2009).

¹⁸Alternative specifications – for example, collapsing the instrument matrix, increasing the number of lags per time period, including different lags – generate equally significant coefficient estimates for both loans and conditions, with acceptable test statistics for overidentification. See the online appendix for a regression with a collapsed instrument matrix, using lags five to ten for loans and lags ten to fifteen for conditions.

¹⁹With the panel dataset, we are limited to using mechanical instruments in GMM, because substantive instruments that significantly predict DPLs exhibit little or no time series variation. Moving to cross-

section data allows us to avoid that problem as well as complications associated with serial correlation in the dependent variable.

²⁰In order to maximize the number of observations, we took 1996 instead of 1995 as the base year.

²¹For instance, the J-test rejects the quadratic specification as the correct model, with a J-statistic of 2.01 with corresponding p-value of 0.046. It does not reject the logarithmic model (J-statistic = -0.61 with p-value 0.54). Similarly, the linear model is rejected in favor of the logarithmic (J-statistic = 2.18, p-value = 0.031), without rejecting the logarithmic model (J-statistic = -0.10 with p-value = 0.981).

²²Data on closing years of policy loans were extracted from a less comprehensive dataset.

²³When the Fraser Institute index is included as the dependent variable, the time period under investigation expands from 1995-2008 to 1980-2008. This might explain the positive time trend in equation 4 of table 4.

²⁴Other specifications generate similar results and are available upon request.

²⁵“Political freedoms” varies from 1 (most democratic) to 7 (least democratic), so a negative coefficient implies that more political freedoms are associated with higher CPIA ratings.

²⁶Numerous studies have demonstrated the positive effects on growth and development of the macroeconomic and structural policies reflected in CPIA clusters A and B. Some of the more widely-cited examples include Barro (1991), Fischer (1993), King and Levine (1993), Sachs and Warner (1995), Frankel and Romer (1999) and Burnside and Dollar (2000).

References

- Agostino, M., 2008. World Bank conditional loans and private investment in recipient countries. *World Development* 36 (10), 1692 – 1708.
- Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. *The Quarterly Journal of Economics* 114 (4), 1243–1284.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. *Journal of Economic Growth* 8 (2), 155–194.
- Alesina, A., Drazen, A., 1991. Why are stabilizations delayed? *The American Economic Review* 81 (5), pp. 1170–1188.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58 (2), 277–97.
- Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1), 29–51.
- Bambaci, J., Saront, T., Tommasi, M., 2002. The political economy of economic reforms in Argentina. *Journal of Policy Reform* 2 (2), 75–88.
- Banerjee, A. V., 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* 107 (3), 797–817.
- Barro, R. J., 1991. Economic growth in a cross section of countries. *The Quarterly Journal of Economics* 106 (2), pp. 407–443.
- Barro, R. J., Lee, J.-W., 2005. IMF programs: Who is chosen and what are the effects? *Journal of Monetary Economics* 52 (7), 1245 – 1269.
- Bazzi, S., Clemens, M. A., 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics* 5 (2), 152–86.
- Beck, T., Clarke, G., Groff, A., Keefer, P., Walsh, P., 2001. New tools in comparative political economy: The database of political institutions. *The World Bank Economic Review* 15 (1), pp. 165–176.
URL <http://go.worldbank.org/2EAGGLRZ40>
- Besley, T., Persson, T., 2011. *Pillars of Prosperity*. Princeton University Press, Princeton and Oxford.
- Bird, G., Rowlands, D., 2001. World Bank lending and other financial flows: Is there a connection? *Journal of Development Studies* 37 (5), 83–103.
- Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1), 115–143.

- Bond, S. R., Hoeffler, A., Temple, J., 2001. GMM estimation of empirical growth models. C.E.P.R. Discussion Papers 3048, Center for Economic Policy Research.
- Boockmann, B., Dreher, A., 2003. The contribution of the IMF and the World Bank to economic freedom. *European Journal of Political Economy* 19 (3), 633 – 649.
- Boone, P., 1996. Politics and the effectiveness of foreign aid. *European Economic Review* 40 (2), 289 – 329.
- Bun, M. J. G., Windmeijer, F., 2010. The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13 (1), 95–126.
- Burnside, C., Dollar, D., 2000. Aid, policies, and growth. *The American Economic Review* 90 (4), 847–868.
- Butkiewicz, J. L., Yanikkaya, H., 2005. The effects of IMF and World Bank lending on long-run economic growth: An empirical analysis. *World Development* 33 (3), 371 – 391.
- Chauvet, L., Collier, P., 2009. Elections and economic policy in developing countries. *Economic Policy* 24, 509–550.
- Clemens, M. A., Radelet, S., Bhavnani, R. R., Bazzi, S., 2012. Counting chickens when they hatch: Timing and the effects of aid on growth. *The Economic Journal* 122 (561), 590–617.
- Collier, P., 2000. Ethnicity, politics and economic performance. *Economics and Politics* 12 (3), 225–245.
- Collier, P., Guillaumont, P., Guillaumont, S., Gunning, J. W., 1997. Redesigning conditionality. *World Development* 25 (9), 1399–1407.
- Djankov, S., Montalvo, J., Reynal-Querol, M., September 2008. The curse of aid. *Journal of Economic Growth* 13 (3), 169–194.
- Dreher, A., Lamla, M. J., Lein, S. M., Somogyi, F., 2009. The impact of political leaders’ profession and education on reforms. *Journal of Comparative Economics* 37 (1), 169–193.
- Dreher, A., Sturm, J.-E., 2012. Do the IMF and the World Bank influence voting in the UN General Assembly? *Public Choice* 151 (1), 363–397.
- Easterly, W., 2005. What did structural adjustment adjust? The association of policies and growth with repeated IMF and World Bank adjustment loans. *Journal of Development Economics* 76, 1–22.
- Fischer, S., 1993. The role of macroeconomic factors in growth. *Journal of Monetary Economics* 32 (3), 485–512.
- Frankel, J. A., Romer, D. H., 1999. Does Trade Cause Growth? *American Economic Review* 89 (3), 379–399.
- Gwartney, J., Lawson, R., Hall, J., 2013. Economic freedom of the World. 2013 annual report. The Fraser

Institute.

- Heckelman, J. C., Knack, S., 2008. Foreign aid and market-liberalizing reform. *Economica* 75, 524–548.
- Heckman, J. J., 1979. Sample selection bias as a specification error. *Econometrica* 47 (1), 153–61.
- IEG, 2013. World Bank Group Assistance to Low-Income Fragile and Conflict-Affected States. World Bank, Washington, D.C.
- Jayarajah, C., Branson, W., 1995. Structural and sectoral adjustment: World Bank experience, 1980-92. World Bank, Washington, D.C.
- Jones, C., Morrissey, O., Nelson, D., 2011. Did the World Bank drive tariff reforms in Eastern Africa? *World Development* 39 (3), 324 – 335.
- Kapur, D., Lewis, J. P., Webb, R., 1997. The World Bank: Its first half century. Volume 1. History. Brookings Institution Press, Washington, D.C.
- Kaufmann, D., Kraay, A., Mastruzzi, M., 2010. The worldwide governance indicators: methodology and analytical issues. Policy Research Working Paper Series 5430, The World Bank.
- Kilby, C., 2005. World Bank lending and regulation. *Economic Systems* 29 (4), 384 – 407.
- Kilby, C., 2011. The political economy of project preparation: An empirical analysis of World Bank projects. Villanova School of Business Department of Economics and Statistics Working Paper Series 14, Villanova School of Business Department of Economics and Statistics.
- Kilby, C., 2013. The political economy of project preparation: An empirical analysis of World Bank projects. *Journal of Development Economics* 105 (C), 211–225.
- Killick, T., Gunatilaka, R., Marr, A., 1998. Aid and the political economy of policy change. Routledge, London and New York.
- King, R. G., Levine, R., 1993. Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics* 108 (3), 717–737.
- Koeberle, S. G., 2003. Should policy-based lending still involve conditionality? *The World Bank Research Observer* 18 (2), 249–273.
- Miller, T., Holmes, K., Feulner, E., 2013. 2013 Index of Economic Freedom. The Heritage Foundation and The Wall Street Journal.
- Moser, C., Sturm, J.-E., 2011. Explaining IMF lending decisions after the Cold War. *The Review of International Organizations* 6 (3), 307–340.
- Mosley, P., Harrigan, J., Toye, J., 1991. Aid and Power: The World Bank and policy-based lending Volume 1. Analysis and policy proposals. Routledge, London and New York.

- Mullainathan, S., 2006. Development economics through the lens of psychology. Proceedings of the Annual Bank Conference on Development Economics.
- OPCS, 2009. Country policy and institutional assessments: 2009 assessment questionnaire. Operations Policy and Country Services, World Bank.
- Pritchett, L., de Weijer, F., 2010. Fragile states: Stuck in a capability trap? Background paper for the 2011 World Development Report.
- Rodrik, D., 1990. How should structural adjustment programs be designed? *World Development* 18 (7), 933 – 947.
- Rodrik, D., 1992. The limits of trade policy reform in developing countries. *Journal of Economic Perspectives* 6 (1), 87–105.
- Rodrik, D., 1996. Understanding economic policy reform. *Journal of Economic Literature* 34 (1), pp. 9–41.
- Rodrik, D., 2006. Goodbye Washington consensus, hello Washington confusion? *Journal of Economic Literature* 44 (4), 973–987.
- Rodrik, D., 2008. Second-best institutions. *The American Economic Review* 98 (2), 100–104.
- Roodman, D., 2007. Macro aid effectiveness research: A guide for the perplexed. Center for Global Development Working Papers 135.
- Roodman, D., 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71 (1), 135–158.
- Rose, A. K., 2006. Size really doesn't matter: In search of a national scale effect. *Journal of the Japanese and International Economies* 20 (4), 482 – 507.
- Sachs, J. D., Warner, A., 1995. Economic reform and the process of global integration. *Brookings Papers on Economic Activity* 26 (1), 1–118.
- Smets, L., Knack, S., Molenaers, N., 2013. Political ideology, quality at entry and the success of economic reform programs. *The Review of International Organizations* 8, 447–476.
- Stiglitz, J., 1998. Towards a new paradigm for development: Strategies, policies and processes. Prebisch Lecture at UNCTAD, Geneva.
- Svensson, J., 2000. When is foreign aid policy credible? Aid dependence and conditionality. *Journal of Development Economics* 61 (1), 61–84.
- Svensson, J., 2003. Why conditional aid does not work and what can be done about it? *Journal of Development Economics* 70 (2), 381–402.
- Williamson, J., 1994. The Political Economy of Policy Reform. No. 68 in Peterson Institute Press: All

Books. Peterson Institute for International Economics.

Wooldridge, J. M., 1995. Score diagnostics for linear models estimated by two stage least squares. In: Maddala, G. S., Phillips, P. C. B., Srinivasan, T. N. (Eds.), *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*. Blackwell, Oxford, pp. 66 – 87.

World Bank, 1989. *Adjustment Lending: An Evaluation of Ten Years of Experience*. World Bank, Washington, D.C.

World Bank, 2004. *2003 Annual Review of Development Effectiveness: The effectiveness of Bank support for policy reform*. World Bank, Washington, D.C.

World Bank, 2006. *Development Policy Retrospective 2006*. World Bank, Washington, D.C.

World Bank, 2009. *Development Policy Retrospective 2009: Flexibility, Customization and Results*. World Bank, Washington, D.C.

World Bank, 2013. *Development Policy Retrospective 2012: Results, Risks, and Reforms*. World Bank, Washington, D.C.